# A Rapid Land Cover Classification Method For Use in Urban Watershed Analysis

Erik Botsford<sup>1</sup>, Kristina Hill<sup>2</sup>, and Derek Booth<sup>3</sup>

<sup>1</sup> Department of Urban Design and Planning

<sup>2</sup> Department of Landscape Architecture

<sup>3</sup> Center for Urban Water Resources Management University of Washington, Seattle WA 98195

#### INTRODUCTION

Modifications of the land surface during urbanization can produce tremendous changes in the patterns and the processes of stormwater runoff. These changes result from clearing vegetation, compacting soil, ditching and draining, and finally covering the land surface with impervious roofs and roads. The infiltration capacity of these covered areas is lowered to zero, and much of the remaining soil-covered area is trampled to a near-impervious state. Compacted, stripped, or paved-over soil also has lower storage volumes, and so even if precipitation can infiltrate, the soil reaches surface saturation more rapidly and more frequently. This results in pervasive changes to water quantity, water quality, and the associated ecological function of streams and riparian areas.

In addition to changes in how rainfall is absorbed or runs off of hillslopes, urbanization affects other elements of the drainage system. Gutters, drains, and storm sewers are laid in the urbanized area to convey runoff rapidly to stream channels. Natural channels are often straightened, deepened, or lined with concrete to make them hydraulically smoother. Each of these changes increases the efficiency of the channel, transmitting the flood wave downstream faster and with less retardation by the channel and destroys the habitat for stream biota.

Because of the profound effect of urban development on aquatic systems, characterizing the land cover of a region is critical for a variety of resource-management applications. In the Pacific Northwest, this characterization has been used most commonly to correlate the intensity of human activity with observed stream or wetland conditions, in order to predict the health of the stream system or to guide the allocation of mitigation efforts. For example, measured biological conditions in lowland streams are regularly presented in terms of "impervious area percentage" of the contributing watershed. Land cover is a primary input parameter for numerical hydrologic models (such as the Hydrologic Simulation Program Fortran [HSPF], widely used by the surface-water management agencies of King County, Snohomish County, the cities of Seattle and Bellevue, and the consultants of these and smaller jurisdictions throughout western Washington). Every one of the \$20+ million in capital projects planned or under construction by King County Water and Land Resources Division, for example, is designed using HSPF with land cover as a primary, determining input.

Unfortunately, there is little consistency or quality control in how land-cover data are collected and analyzed. Some of this variety is entirely appropriate—the methods and the products for assessing wilderness-area potential in the Cascade Range have little overlap with those used to plan optimal siting of commuter-rail stations, for example.

Yet certain applications constantly reemerge, and so typical procedures have been developed but only on an *ad hoc* basis.

The characterization of land cover for purposes of evaluating and assessing aquatic-system conditions is one such application. Yet the imprecision of the methodology currently used to classify land cover belies the significance of the results: typically, recent 1:12,000-scale airphotos (within the last 2-3 years) are manually discriminated by a technician into eight or so different "classes," of which four discriminate urban development of different densities and the remainder characterize the unbuilt areas. Discrimination is at the judgement of the operator, following established guidelines; typical minimum unit areas are one to five acres (about 100 m minimum dimensions); and subsequent ground truthing is nonexistent. Typical analyses require about 1 person-week for a 10 mi² area, and once the operator is trained there are no opportunities for greater speed—every new area requires an equivalent level of effort. *This* is the procedure against which any alternative method should be compared.

Remotely sensed data from satellites provide an alternative source of information on land cover over very large areas. The traditional approach to classifying remotely sensed data from satellites into discrete classes of land cover involves a lengthy process of automated classification, clustering of spectral signatures, much fine-tuning, and an eventual supervised classification. This process can be both time and resource-intensive. It is also continually being refined, and so the methodologies are not consistent.

We have developed an alternative approach using Landsat satellite imagery to produce the same general type of land-cover characterization as has currently found widespread acceptance and use across the region. However, our methodology does so in a way that achieves maximum utility and consistency for a particular group of users—individuals and agencies needing to assess watershed conditions in the urban, and urbanizing, parts of western Washington. The classes of land cover produced have been chosen to reflect the categories that can be readily distinguished in the satellite data and to have important differences in their associated runoff and watershed characteristics.

The advantages of such an approach are obvious. The algorithm is developed only once; after completion, it can be applied rapidly to any other selected area through GIS software. It does not depend on the discretion of individual operators and so the results are reproducible. These advantages have not been lost on public agencies, but those agencies are not equipped to pursue such efforts systematically, given project-related geographic boundaries, time constraints, staff turnover, and the difficulty of interagency communication. With suitable testing and documentation, the release of these data layers through the University of Washington may encourage agencies across the region to adopt a uniform methodology, resulting in a degree of uniformity in data collection, analysis, and reporting of these data that is currently unavailable.

#### **METHODOLOGY**

The methodology used in this project is summarized in Figure 1. The area of this analysis was a portion of the Puget Sound lowlands of northwestern Washington State. The area extends from the city of Olympia in the south to Everett in the north, and it includes the entire Seattle-Tacoma-Everett metropolitan area from Puget Sound east to the foothills of the Cascade Mountain range. The study site was chosen to cover a broad range of urban, suburban, and rural areas while excluding those areas with extreme topographic relief and little or no urban development.

Our classification scheme followed a multi-step process that was designed to be intuitive while yielding accurate results. It consisted of:

- 1. Combination and manipulation of the raw satellite images;
- 2. Selection of training sites, where different land-cover categories could be defined;
- 3. Extraction of the "typical" Landsat signatures for each coverage;
- 4. Classification of the entire image, following the characteristics defined for each class; and
- 5. Assessment of the classification's accuracy by checking actual field conditions at selected locations.

## **STEP 1: Image Manipulation**

Landsat satellite images from 1991 and 1998 were obtained for northwestern Washington State. The resolution of these images was 30 meters, meaning that each pixel represented an area of 900 m<sup>2</sup>. The raw Landsat images were imported into the ERDAS Imagine software package and geocorrected to the UTM projection (zone 10N, spheroid Clarke 1866, datum NAD27) utilizing digital orthophotos (DOQQ's) as the reference projection source. This allowed the raw Landsat and classified images to be compared directly to the DOQQ's using the geographic linking function in the Imagine viewer.

The raw Landsat images contain 7 layers, or bands, that correspond to reflectances received by the satellite sensors in various wavelength ranges. Bands 2, 3 and 4 (corresponding to the visible blue, near infrared, and mid-range infrared portions of the EM spectrum) were extracted from the raw image. These bands were selected for their ability to discriminate and detect characteristics of vegetation such as chlorophyll absorption, biomass content, and vegetation vigor (Lillesand and Kiefer 1994). Using Imagine's Model Maker module, bands 3 and 4 were used to create a Normalized Difference Vegetation Index (NDVI) layer. The NDVI layer is particularly useful in discerning vegetation vigor and subtler variations in vegetative land cover and is commonly used in remote sensing research. The NDVI layer was then combined with the individual bands 2, 3 and 4 using the STACKLAYER function to yield one composite image with 4 layers. This image is hereafter referred to as the "234N image." The 234N image was clipped to the extent of the study area and the remainder of the image was discarded.

## **STEP 2: Training Sites**

The next step in the process was the identification of and delineation of training sites used to define the characteristic pattern, or "signature," in the 234N image for each land-cover category. "Training sites" are areas of known land cover, usually no more than 1000 m² in size, determined from ground truthing in the field or from inspection of digital orthophoto quarter quadrangles (DOQQ's). We used a combination of both methods to obtain a series of suitable training sites for each desired class. Training sites were digitized onto a separate layer in Imagine that could be overlaid onto the Landsat and DOQQ images.

Separate training sites were created for a total of 9 classes, using a two-tier scheme (Table 1). The top scheme, called *Toplevel*, consisted of 4 broad land cover classes: "intense urban" (land nearly completely paved or built upon), "water," "vegetation," and "broad urban" (essentially everything remaining). The second tier, called *5Veg*, consisted of 5 finer classes that subdivided the "vegetation" and "broad urban" classes of Toplevel into "deciduous vegetation," "coniferous vegetation," "grassy/shrubby vegetation," "forested urban" (developed land with significant canopy coverage), and "grassy urban" (developed land with few trees but significant grass coverage).

## Toplevel

- Intense urban
- Water
- Vegetation
- Broad urban

## 5Veg

- Deciduous vegetation
- Coniferous vegetation
- Grass/shrub vegetation
- Forested urban
- Grassy urban

Table 1. Classes used in the land-cover classification.

## **STEP 3: Signature Extraction**

Once selected, the outlines of the training sites were overlaid on the 234N image. Using the Signature Editor module in Imagine, spectral signatures were simultaneously extracted from each of the 4 layers of the 234N image, yielding "signatures" for each training site within each class. Signatures correspond to a cluster of reflectance values

within each band. Signatures within each class were then combined to obtain a single spectral signature range in 4-dimensional space for each class. The output of this step thus consisted of 9 distinct sets of signatures.

## **STEP 4: Supervised Classification**

The first supervised classification was conducted on the 234N image using the Toplevel, or first-tier, signatures (i.e. intense urban, water, vegetation, and broad urban). The classification utilized the parallelepiped non-parametric rule and the maximumlikelihood parametric rules in classification of individual pixels of the Landsat image. Remote sensing software packages, such as Imagine, construct bounding boxes, or parallelpipeds, around clusters of signatures collected from the training sites. The limits of these parallelpipeds represent each individual class in multi-spectral space. In classifying the entire image, certain decision rules govern how pixels are classified (see chapter 7 in Lillesand and Kiefer [1994] for a more thorough explanation of classification decision rules). Pixels whose multi-spectral reflectance values fall within the limits of a parallelpiped are immediately classified to that parallelpiped's corresponding class. In areas where parallepipeds overlap or for pixels that fall outside parallepiped limits, the maximum likelihood decision rule determines the classification. This rule calculates the statistical probability of a pixel belonging to a particular class, based on the variance and covariance of the spectral signatures. The combination of the parallepiped and maximum-likelihood decision rules results in an output map in which no pixels are left unclassified.

The output from this process, termed the "Toplevel image," consisted of 4 classes corresponding to the 4 Toplevel signatures in Table 1. This image was then recoded, using the RECODE function in Imagine, to create masks for each of the 4 classes. From this process, we created 4 masks, one for each Toplevel category. Masks only contain pixel values of 0 and 1 and are useful in cropping other maps.

A second supervised classification was then conducted on the 234N image using the second-tier signatures (deciduous vegetation, coniferous vegetation, grass/shrub, forested urban, and grassy urban). This classification used the same parallelepiped-maximum likelihood classification rules and produced a classified image with 5 classes corresponding to the 5Veg signatures in Table 1. This image is referred to as the "5Veg image."

This new five-class image took all pixels in the study area and assigned them to one of the 5Veg categories. Some misclassification occurred during this step as areas of water or intense urban cover were classified to one of the 5Veg classes. To eliminate these misclassified pixels, we used the Toplevel broad urban mask to crop the 5Veg image to the areas not originally classified as water, vegetation, or intense urban. The resulting image contained a 5-class classified image of the broad urban study area. We then used the Toplevel vegetation mask to crop the original 5Veg image to the areas that were classified as vegetation in the Toplevel classification. This step allowed us to break the coarse Toplevel classes of vegetation and broad urban into the finer 5Veg classes.

At this point we had four images: a single-class "intense urban" image, a single-class "water" image, and two images with five classes for the vegetated and broad urban areas. Each of the images was mutually exclusive, with their own unique coverage within the study area. The final step was to combine each of these images into a final composite seven-class image (Table 2). Each of the four images was recoded using the RECODE function in Imagine to give each class a unique value. All the images were then combined to obtain a final image called "Final-classed".

#### **Final Classes**

Urban
Water
Deciduous vegetation
Coniferous vegetation
Grass/shrub
Forested urban
Grassy urban

Table 2. Final classes used in classification.

## **STEP 5: Accuracy Assessment**

The final step involves an accuracy assessment of the classed images for 1991 and 1998. This work is in progress as of mid-1999 under the supervision of Derek Booth and Kristina Hill. It entails a pixel-by-pixel error check using the classified images and some source of ground truthing (either DOQQ's or in-field truthing). An error check that examines a larger array of pixels simultaneously will also be used to determine the aggregate accuracy of the images.

#### PRELIMINARY RESULTS

We have conducted a qualitative evaluation of the classification outputs using digital and printed overlays of the classified images and the DOQQ's. The printed results of this evaluation are presented in Figures 2-7. This evaluation suggests that our classification method will be sufficiently accurate for use in land cover-related and urban hydrological research, particularly in comparison to the slow and inconsistent methods currently in use throughout the region. Where aggregated across a watershed, the classification seems to be fairly accurate in discriminating between the 7 classes of land cover used. The classification scheme also seems to be able to successfully discern between forested urban and grassy urban areas (see figure X). While we have not explicitly determined a threshold of tree cover or grass cover that defines each of these classes, a visual examination of the overlays seems to indicate that our scheme matches the categories well.

#### **FURTHER APPLICATIONS**

Although multiple land-cover catagories have great utility, there is great appeal to identifying a single "index" variable that characterizes the magnitude of urban development in a watershed. Patterns can be readily displayed, correlations are simplified, and communication between scientists and planners is enhanced. Yet urban development comes in many styles, occurs on many different types of landscapes, and is accompanied by a variety of mitigation measures designed to reduce its negative consequences on downstream watercourses. So any simple correlation between any single measure of urbanization and aquatic-system condition are unlikely to be precise.

Past efforts to quantify the degree of urban development have not been consistent. Recent and historical use of the most widely accepted parameter, percent impervious area in the contributing watershed, has been carefully documented in a recent review article (Schueler, 1995) but several issues remain ambiguous. Most significant of these is the distinction between *total impervious area* (TIA) and *effective impervious area* (EIA).

TIA is the "intuitive" definition of imperviousness: that fraction of the watershed covered by constructed, non-infiltrating surfaces such as concrete, asphalt, and buildings. Hydrologically this definition is incomplete for two reasons. First, it ignores nominally "pervious" surfaces that are sufficiently compacted or otherwise so low in permeability that the rate of runoff from them are similar or indistinguishable from pavement. For example, Wigmosta and others (1994) found that the impervious unitarea runoff was only 20 percent greater than that from pervious areas, primarily thin sodded lawns over glacial till, in a western Washington residential subdivision. Clearly, this hydrologic contribution cannot be ignored entirely.

The second limitation of TIA is that it includes some paved surfaces that may contribute nothing to the storm-runoff response of the downstream channel. A gazebo in the middle of parkland, for example, probably will impose no hydrologic changes into the watershed except a very localized elevation of soil moisture at the drip line of its roof. Less obvious, but still relevant, will be the different downstream consequences of rooftops that drain alternatively into a piped storm-drain system, with direct discharge into a natural stream, or onto splashblocks that disperse the runoff onto the garden at each corner of the building.

The first of these TIA limitations, the production of significant runoff from nominally pervious surfaces, is typically ignored in the characterization of urban development. The reason for such an approach lies in the difficulty in identifying such areas and estimating their contribution, although site-specific studies demonstrate that these tasks can be accomplished with simple field methods and the resulting hydrologic insights are often valuable (Burges and others, 1989). Furthermore, the degree to which pervious areas shed water as overland flow should be related, albeit imperfectly, with the amount of impervious area: where construction and development is more intense and covers progressively greater fractions of the watershed, the more likely that the intervening green spaces have been stripped and compacted during construction and only imperfectly rehabilitated for their hydrologic functions during subsequent "landscaping."

The second of these TIA limitations, inclusion of non-contributing impervious areas, is formally addressed through the concept of effective impervious areas, defined as the

impervious surfaces with direct hydraulic connection to the downstream drainage (or stream) system. Thus any part of the TIA that drains onto pervious (i.e. "green") ground is excluded from the measurement of EIA. This parameter, at least conceptually, captures the hydrologic significance of imperviousness. EIA is the parameter normally used to characterize urban development in hydrologic models.

Yet the direct measurement of EIA is complicated. Studies designed specifically to quantify this parameter must make direct, independent measurements of both TIA and EIA (Alley and Veenhuis, 1983; Laenen, 1983; Prysch and Ebbert, 1986). The results can then be generalized either as either a correlation between the two parameters or as a "typical" value for a given land use. Alley and Veenhuis found that [EIA] = 0.15  $[TIA]^{1.41}$  in their highly urbanized watersheds in Denver, Colorado ( $r^2 = 0.98$ ). Using the other approach (I,e. typical land-use values), Dinicola (1989) compiled the findings of these earlier studies to recommend a single set of impervious-area values based on five land-use categories for use in studies of western Washington watersheds (Table 3).

LAND USE	TIA (%)	EIA (%)
Low density residential (1 unit per 2-5	10	4
acres)		
Medium density residential (1 unit per acre)	20	10
"Suburban" density (4 units per acre)	35	24
High density (multi-family or 8+ units per	60	48
acre)		
Commercial and industrial	90	86

Table 3. Presumed Relationship between Imperviousness and Land Use (from Dinicola, 1989).

Because our analysis is being conducted at a much finer scale (30-m pixels) and detects only land-cover differences, we can evaluate only *total* imperviousness. Landuse categories, and thus EIA, might be inferred from larger clusters and patterns of individual pixels, but this lies outside the scope of this present effort. Based on detailed measurements of impervious areas oat our training sites, we anticipate having median and ranges of TIA values associated with each of the nine land-cover categories at the conclusion of this project.

#### REFERENCES

Alley, W. A., and Veenhuis, J. E., 1983. Effective impervious area in urban runoff modeling. Journal of Hydrological Engineering, ASCE, 109(2):313-319.

Burges, S. J., Stoker, B. A., Wigmosta, M. S., and Moeller, R. A., 1989. Hydrological information and analyses required for mitigating hydrologic effects of urbanization. University of Washington, Department of Civil Engineering, Water Resources Series Technical Report no. 117, 131 pp.

- Dinicola, R. S., 1989. Characterization and simulation of rainfall-runoff relations for headwater basins in western King and Snohomish Counties, Washington state. U. S. Geological Survey Water-Resources Investigation Report 89-4052, 52 pp.
- Laenen, A., 1983. Storm runoff as related to urbanization based on data collected in Salem and Portland, and generalized for the Willamette Valley, Oregon. U. S. Geological Survey Water-Resources Investigations report 83-4238, 9 pp.

#### Lillesand and Kiefer 1994

- Prysch, E. A., and Ebbert, J. C., 1986. Quantity and quality of storm runoff from three urban catchments in Bellevue, Washington. U. S. Geological Survey Water-Resources Investigations Report 86-4000, 85 pp.
- Schueler, T., 1995. The importance of imperviousness. Watershed Protection Techniques 1(3):100-111.
- Wigmosta, M. S., Burges, S. J., and Meena, J. M., 1994. Modeling and monitoring to predict spatial and temporal hydrologic characteristics in small catchments. Report to U. S. Geological Survey, University of Washington Water Resources Series Technical Report No. 137, 223 pp.

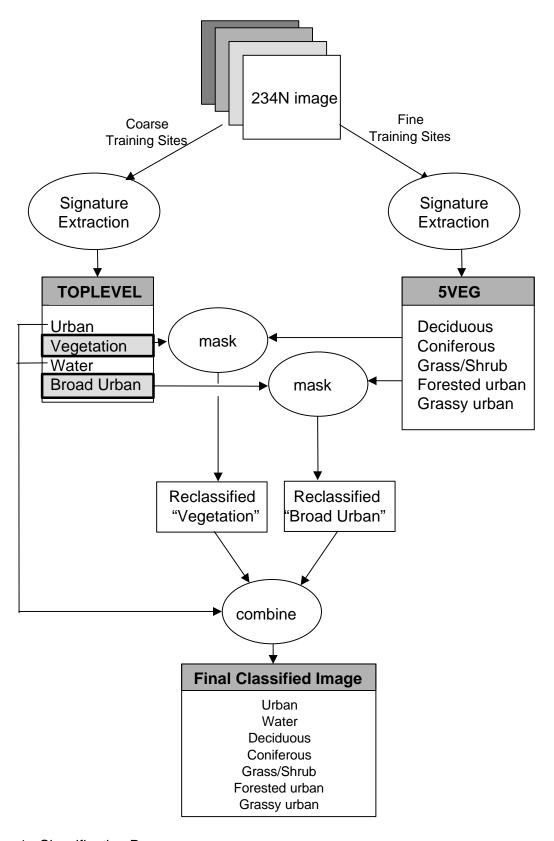


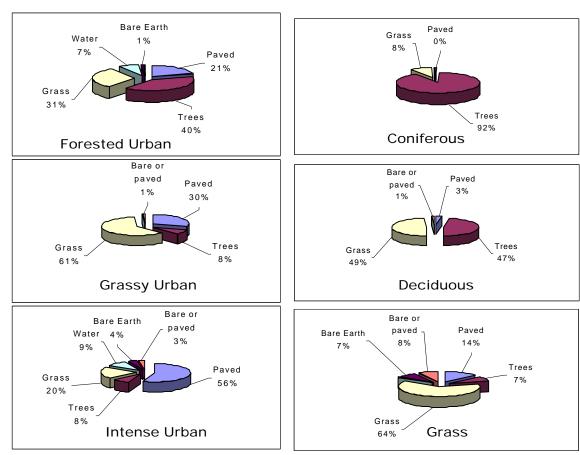
Figure 1. Classification Process

## APPENDIX—ERROR-CHECKING THE CLASSIFIED 1991 IMAGE

After classifying the entire 1991 Landsat image, randomly selected classified pixels were compared with low-elevation orthophotos to determine the actual land cover that corresponds to each category in the classified image. The categories, and their corresponding average land cover percentages, are as follows:

_	Actual Land Cover from Orthophotos (percentages, averaged for 100 pixels)			
	Open water	Trees	Shrubs/ grass	Pavement or bare earth
Categories from the classified Landsat image:				
"UNDEVELOPED"				
Open water	100	0	0	0
Coniferous vegetation		91	8	1
Deciduous vegetation		47	49	4
"DEVELOPED"				
Grassy/shrubby vegetation		8	63	29
Forested urban	7	39	31	23
Grassy urban		8	61	31
Intense urban	9	8	21	62

## 1991 LAND COVER CATEGORIES



Note: See the associated spreadsheet for the percentage of land-cover values for *each* of the 100 pixels evaluated for each classification category. The category "Water" had 100% observed water coverage and so is not graphed here.